PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/70, C07K 13/00, 7/00 C12P 21/02, C12N 15/62, 1/21 // (C12N 1/21, C12R 1/19)	A1	(11) International Publication Number: WO 91/07496
		(43) International Publication Date: 30 May 1991 (30.05.91)

(21) International Application Number: PCT/US90/06354

(22) International Filing Date: 1 November 1990 (01.11.90)

(30) Priority data:
433,936 9 November 1989 (09.11.89) US

(71) Applicant: ALLIED-SIGNAL INC. [US/US]; Law Department (C.A. McNally), P.O. Box 2245R, Morristown, NJ 07962-2245 (US).

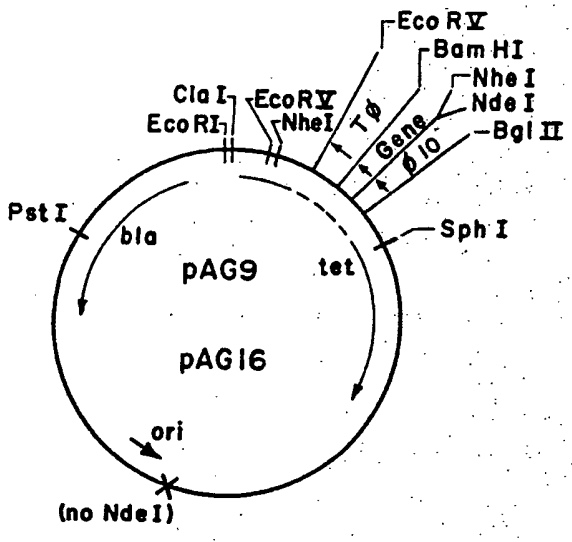
(72) Inventors: SALERNO, Anthony, Joseph ; 19 Francine Place, Millington, NJ 07946 (US). GOLDBERG, Ina ; 100 Sunrise Drive, Gillette, NJ 07933 (US).

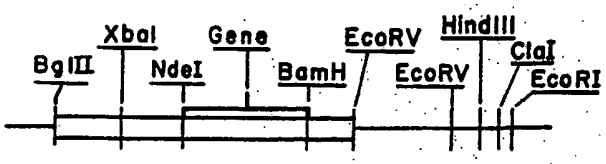
(74) Agent: ROONEY, Gerard, P.; Allied-Signal Inc., Law Department (C.A. McNally), P.O. Box 2245R, Morristown, NJ 07962-2245 (US).

(81) Designated States: AT (European patent), BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: CLONING AND EXPRESSION OF GENES ENCODING FOR POLYPEPTIDES COMPRISED OF ONE OR MORE REPEATING AMINO ACID SEQUENCES





(57) Abstract

Cloning and expression genes coding for polypeptides comprised of one or more repeating amino acid sequences.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	ML	Mali
AU	Australia	FR	France	MN	Mongolia
BB	Barbados	GA	Gabon	MR	Mauritania
BE	Belgium	GB	United Kingdom	MW	Malawi
BF	Burkina Faso	GN	Guinea	NL	Netherlands
BG	Bulgaria	GR	Greece	NO	Norway
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	SD	Sudan
CF	Central African Republic	KP	Democratic People's Republic of Korea	SE	Sweden
CG	Congo			SN	Senegal
CH	Switzerland	KR	Republic of Korea	SU	Soviet Union
CI	Côte d'Ivoire	LI	Liechtenstein	TD	Chad
CM	Cameroon	LK	Sri Lanka	TG	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		
ES	Spain	MG	Madagascar		

CLONING AND EXPRESSION OF GENES ENCODING FOR POLYPEPTIDES COMPRISED OF ONE OR MORE REPEATING AMINO ACID SEQUENCES

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates to cloning and expression of genes encoding for polypeptides comprised of one or more repeating amino acid sequences, to polypeptide products resulting from such cloning and expression and to transformed microbes for use in such production. Another aspect of this invention relates to processes for genetically engineering such microbes and to plasmids and vectors for use in such engineering.

2. Prior Art

Various p9polypeptides having recurring amino acid sequences exhibit beneficial properties. Illustrative of such polypeptides are those of the collagen family such as collagen, elastin, fibronectin, lamenin and other fibrous proteins, and structural proteins such as annelid or arthoopod silks, bacterial flagellin, resilin, eucaryotic egg shell proteins, insect cuticle proteins and architectural proteins involved with eucaryotic development processes such as tissue organization.

Still other useful polypeptides having recurring amino acid sequences include adhesive substances secreted by purple shellfish (mussel) such as mussel of the genus Mytilis and other spineless animals in the sea for attachment of their bodies to underwater structural materials. For example, the purple shellfish (Mytilis edulis) secretes from its foot an adhesive substance which hardens subsequently to bond permanently to the substrate. The main component of the bonding plate which M. edulis puts out is a hydroxylated proteins of about 130,000 dallons having recurring dicopeptide units. See United States Patent Nos. 4,496,397; 4,585,585; 4,687,740;

and 4,808,702 and J.H. White, J. Biol. Chem., 258, 2811-2915 (1983).

Various methods have been heretofore used to form or otherwise obtain useful polypeptides having recurring amino acid sequences, all of which suffer from a number of inherent disadvantages. For example, in one prior art method, the polypeptide is formed using solid-phase chemical synthesis of the repeating peptide sequences. This procedure however, is labor intensive and limited in the quantity of polypeptides that can be produced. It is therefore not feasible to use this method in the large scale commercial production of useful polypeptides.

Other polypeptides such as polyphenolic proteins from the mussel genus Mytilis is isolated as the natural occurring polypeptide from natural products. This procedure is also later intensive and provides a limited quantity of the desired product. It is not practical to use this procedure in the production of polypeptides.

Microbial production of polypeptides having recucrring peptide sequences offers several advantages. For example, microbial production methods provides large quantity production, sheap production and timely production.

Procedures for genetically engineering microbes to produce polypeptides are known. Illustrative of certain aspects of these procedures relevant to this application are those described in G.D. Stormo, T.D. Scheider and L.M. Gold, Nucleic Acids Research 10, 297-2996 (1982); A. Shatzman, Y.S. Ho and M. Rosenberg in Experimental Manipulation of Gene Expression, pp. 1-14. M. Inouye, ed. (Academic Press, 1983); A. Rattray, S. Altuvia, G. Mahagna, A.B. Oppenheim and M. Gottesman, Journal of Bacteriology 159, 238-242 (1984). Modern biochemical advances in genetic technology have led to the introduction of new techniques for transferring genes between species. Many of these techniques are based on the use of plasmid vectors with microorganisms as hosts. These vectors allow establishment and expression of

foreign genes in microorganisms such as bacteria under controllable conditions. See J.G. Sutcliffe and F.M. Ausubel in Genetic Engineering, pp. 83-111. A.M. Chakrabarty, ed. (CRC Press, 1978) and R. Wu, L.-H. Guo and R.C. Scarpella in Genetic Engineering Techniques, pp. 3-21. P.C. Huang, T.T. Kuo and R. Wu, eds. (Academic Press, 1982). A large number of plasmids are now available that allow cloning of either genes with their nautrally associated regulatory DNA sequences or genes which function under the control of regulatory DNA sequences or genes which function under the control of regulatory DNA sequences inherent to the parent plasmid. Many of these plasmids have been applied to the isolation, characterization and expression of many genes, gene fragments or gene promoter sequences. Most of the genes which have been cloned and expressed from plasmid vectors in bacteria such as the gram-negative bacterium Escherichia coli code for proteins which are enzymes or which have a physiologic function (e.g., hormones, blood factors, cell growth factors, etc.). Relatively few genes or gene fragments have been cloned that code for all or part of a structural protein such as components of the extracellular matrix in multicellular higher organisms; these proteins include the collagen family, elastin, fibronectin, laminin and other fibrous proteins. Other structural proteins with interesting physical or chemical properties include the protein or glycoprotein elements of thick, intermediate or this filaments in higher organisms, the annelid or arthropod silks, bacterial flagellin, resilin, eucaryotic egg shell proteins, insect cuticle proteins and architectural proteins involved with eucaryotic developmental processes such as tissue organization. Very few of these cloned genes have been expressed and their protein products isolated, purified and/or biochemically analyzed following their expression in a heterologous bacterial host.

Researches in recombinant DNA technology using the bacterial host E. coli who have been or who are interested

in optimizing foreign gene expression from plasmid vectors have utilized various strategies for increasing protein production from the foreign genes. These strategies include use of runaway replication of the plasmid vector, thermal or chemical induction of the promoter DNA sequence controlling expression of the foreign gene, or use of highly active promoter sequences such as the lac, trp or lpp promoters endogeneous to E. coli or natural or synthetic mutant forms thereof. For illustrative examples of such efforts, see B. Uhlin, S. Molin, P. Gustafsson and K. Nordstrom, Gene 6, 91-106 (1979); K. Backman and M. Ptashne, Cell 13, 65-71 (1978); K. Mordstrom, S. Molin and J. Light, Plasmid 12, 71-90 (1984); and P. Stanssens, E. Remaut and W. Fiers, Gene 36, 211-223 (1985). Hybrid promoters which advantageously use a -35 consensus sequence and a 5' flanking region from one promoter and a portion of a promoter/operator sequence including a -10 region sequence and a Shine-Selgarno sequence from a second natural or synthetic promoter/operator DNA sequence have proven particularly useful for high level expression of foreign genes, in E. coli. See literature, in the case of hybrid trp-lac promoters, such as H.A. DeBoer, L.J. Comstock and M. Vasser, Proc. Natl. Acad. Sci. 80, 21-25 (1983); E. Amann, J. Brosius and M. Ptashne, Gene 25, 167-178 (1983); U.S. Patent 4,551,433 issued Nov. 5, 1985 to H.A. DeBoer, European Patent application 0136090 (filed Aug. 24, 1985) by R. Arentzen and S.R. Petteway, Jr. Plasmid vectors utilizing the controlling elements of the bacteriophage lambda P_L promoter in concert with additional elements such as temperature-sensitive expression of the cI repressor protein governing activity from the P_L promoter and the nutL locus for antitermination activity mediated by the bacteriophage N protein have also provided high levels of foreign gene expression in E. coli and proved comparatively to be as strong or stronger than other strong promoters such as the lacUV5 promoter in E. coli; E. Remaut, P. Stanssens and W. Fiers, Gene 15, 81-93 (1981); U.S. Patent 4,578,355,

issued Mar. 25, 1986 to M. Rosenberg; J.A. Lautenberger, D. Court and T.S. Papas, Gene 23, 75-84 (1983); and European patent application 0131843 (filed Mar. 7, 1984) by H. Aviv, M. Corecki, A. Lavanon, A. Oppenheim, T. Vogel, E. Zeelon and M. Zeevi; and C.A. Caulcott and M. Rhodes, Trends in Biotechnology 4, 142-146 (1986). Most of these publications describe cloning of foreign genes in phase with an initiation condon ATG and production of a fusion protein under the control of the lambda PLOL promoter/operator system, N protein-nutL interaction and the lambda cII gene ribosomal binding site. The product fusion protein then includes some portion of the amino terminus peptide sequence from the bacteriophage lambda cII protein.

Applicants are aware that the Department of Health and Human Services, U.S.A., under the names of T.S. Papas and J.A. Lautenberger filed a U.S. Patent application under Serial No. 6-511,108 on July 6, 1983, covering the plasmid pJL6. Portions of this application have been obtained from the national Technical Information Service, U.S. Department of Commerce. However, the claims are not available and are maintained in confidence. The available portions of the application have been reviewed. The construction of pJL6 is described and its use as a cloning and expression vector for heterologous genes is discussed with relevant examples drawn exclusively from molecular cloning experiments with oncogenes. No mention is made in the available application portions of the use of recombination deficient bacterial hosts, the cloning of synthetic genes or genes coding for structural proteins, or cloning into restriction enzyme recognition sites in pJL6 other than te claI site or the claI-BamHl site pair. All heterologous genes therefore cloned in pJL6 will necessarily produce fusion protein products whereby the foreign gene product cannot be prepared free of amino acid residues on the amino terminus which derive from the lambda cII gene.

A. Seth, P. Lapis, G.F. Van de Woude and T.S. Papas in Gene 42, 49-57 (1986) describe modification of the expression vector pJL6 to yield a class of plasmid vectors which contain in 5' to 3' order: the lambda bacteriophage PLOL promote/operator sequence, an N gene-cro gene fusion polypeptide, the N gene utilization site (nutL), a ribosomal binding site from the lambda cII gene and a restriction enzyme recognition site which is adjacent to the initiation condon ATG and which allows insertion of foreign genes in phase with the initiation codon so as to code for a protein product with at most one extraneous amino acid residue. The plasmids constructed by A. Seth et al. were specifically designed to be cleaved by an appropriate restriction enzyme and treated with S1 nuclease and also have an NdeI restriction site downstream of the unique HpaI, BamHI or KpnI restriction sites described as useful for cloning foreign genes. This article makes no mention of cloning synthetic genes or production of structural proteins for other than the purpose of biochemical research studies. Any advantages of the use of E. coli recombination deficient bacterial hosts for these plasmids is also not disclosed nor discussed by these authors.

H. Aviv et al. (op. cit.) claim as a composition of matter vectors which include in 5' to 3' order: a DNA sequence which contains the promoter and operator P_LO_L from bacteriophage lambda, the N gene utilization sie for binding antiterminator N protein produced by the host cell, a DNA sequence which contains a ribosomal binding site for rendering the mRNA of the desired gene capable of binding to ribosomes within the host cell, an ATG initiation codon or a DNA sequence which is converted into an ATG initiation codon upon insertion of the desired foreign gene into the vector, and a restriction enzyme recognition site for inserting the desired foreign gene into the vector in phase with the ATG initiation codon. This type of vector does not necessarily suffer from potential disadvantages of producing fusion proteins with

unwanted amino acid residues at the amino terminus which canot be conveniently removed. No mention is made in this patent application of cloning of synthetic genes or of genes with repeating amino acid sequences, of cloning of structural proteins or proteins wity interesting physical properties, or of the utility or preferred use of E. coli recombination-deficient bacterial hosts for gene expression from the claimed plasmid vectors.

Gene fusions and hybrid genes have been known in the art of molecular genetics for a number of years. For example, see L. Guarente in Genetic Engineering, Principles and Methods - Volume 6, pp. 233-248 (J.K. Setlow and A. Hollaender, eds.; Plenum Press, 1984) and J.H. Kelly and G.J. Darlington, Annual Reviews of Genetics 19, 273-296 (1985) for reviews. Also see world patent applications Wo 83/03547 (U.S.A. priority date April 14, 1982) by J.L. Little and R.A. Lerner, WO 85/02611 (filed December 12, 1984) by R.A. Houghten for the ScrippsClinic and Research Foundation and WO 86/01210 (filed August 1985) by D.A. Carson , G. Rhodes and R. Houghten for the Scripps Clinic and Research Foundation, and European patent applications EPA 0141484 (GB priority date June 10, 1983) by C. Weissan and H. Weber for Biogen N.V., EPA 0152736 (GB priority date November 1, 1984) by H. Ferres, R.A.G. Smith and A.J. Garman for Beecham Group P.L.C.1, and EPA 0161937 (GB priority date May 16, 1984) by K. Nagai and H.C. Thogerson for Celltech Ltd. All of these patent applications describe the production of fusion or hybrid proteins for a variety of pharmacological agents, enzyme conjugates and diagnostic methods and kits. None of these applications, however, refers to the production of proteins preferred for their physical or structural properties, the production of peptides or proteins from synthetic genes or discusses a requirement to produce recombinant products in recombination-deficient bacterial hosts. Some of these applications claim peptide or protein products with internally repeating amino acid sequ nces, including

oligomers of a native protein, but without exception these products as discussed in the relevant applications are pharmocologically or antigenically active compounds.

As another aspect of the art of molecular cloning pertinent to the invention described herein, it should be noted that several research groups have successfully cloned synthetic genes. Very few of these cloning efforts have focused on peptide or protein products with internally repeating amino acid sequences. The cloning of a synthetic gene coding for a polymeric form of an oligopeptide, specifically the dipeptide L-aspartyl-L-phenylalanine, is disclosed in M.T. Doel et al., Nucleic Acids Research 8, 4575-4592 (1980). A requirement therein for the use of recombination-deficient host is recognized by the employment of E. coli strain HB101 (genotype recA13) which is widely used in the art of molecular cloning. However, these researchers only describe a process for producing polymeric forms of short oligopeptides which could be subsequently broken down chemically or enzymatically into short oligopeptides and do not address any potential advantages to production and use of the polymeric peptides directly. The method described in this reference also is limited to those synthetic genes which can be constructed by annealing two completely complementary oligodeoxynucleotides so as to create DNA hybrids with staggered ends that can further anneal into large oligomeric synthetic DNA sequences. There is no disclosure in this reference of any method to further oligomerize the synthetic gene products into even larger synthetic genes.

Other literature in the art of molecular cloning and peptide or protein expression has dealt with the problem of DNA segment oligomerization. Strategies have been presented in several of these references for specifically and efficiently linking equivalent DNA segments into long DNA sequences which code in an uninterrupted fashion for a large peptide of protein product with internally repeating sequence. See J.L. Hartley and T.J. Gregori, Gene 13,

347-353 (1981); T.A. Willson et al., Gene Analytical Techniques 2, 77-82 (1985); and T. Kempe et al., Gene 39, 239-245 (1985). In contrast to the current invention, none of these references discloses production of synthetic genes coding for repeating amino acid sequences which are of essential value in the polymerized state or discloses the preferred use of recombination-deficient bacterial hosts for plasmid expression vectors bearing synthetic genes. The examples and discussion in these articles bear only on aggregates or oligomers of protein or peptide products which are pharmacologically active or have an undisclosed activity.

U.K. patent application GB 2162190 (filed July 8, 1985) describes a method of producing polypeptide products which are components of silk including those wherein the silk protein comprises sets of the sequence (Gly-Ala-Gyl-Ala-Gyl-Ser).

SUMMARY OF THE INVENTION

One aspect of this invention relates to replicons capable of expressing a polypeptide comprising one or more repeating peptide sequences, said replicon comprising in sequence

an expresion system comprising a promoter, a ribosome binding site and initiation codon; and

one or more structural genes coding for said polypeptide downstream of said epression system, said gene being controllable by said system whereby said genes are expressible to form said polypeptide when said replicon is cloned into a suitable host microbial organism such that the yield of said polypectide is equal to or greater than about 10% by weight bared on the total weight of cellular protein.

Yet another aspect of this invention relates to a method of transforming a microbial organism capable of producing polypeptides comprising one or more repeating peptide sequences, said method comprising the step of

transforming a microbial organism with the replicon of this invention, and relates to microbial organisms resulting from such transformation.

Yet another aspect of this invention relates to a method of producing a polypeptide comprising one or more repeating peptide sequences, said method comprising:

a) growing the transformed microbial organism of this invention in a cellular medium to effect expression of said genes containing the coding sequences for said polypeptides to form said polypeptides;

b) isolating from said microbial organism a fraction comprising said polypeptide and

c) purifying said fraction ro provide said polypeptide.

This invention provides one or more advantages over known replicons, microbial organism and methods. For example, microbial organisms transformed in accordance with this invention provide relatively higher yields of the polypeptide, exhibit enhanced stability and in the preferred embodiments produces un-fused polypeptide products.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be more fully understood and further advantages will become apparent when reference is made to the following details description of the invention and the accompanying drawings in which

FIG. 1 is a physical map of the replicon of this invention.

FIG. 2 is a physical map of the replicons pAG9 and pAG 16.

FIG. 3 is a physical map of the plasma vector pET-3a.

FIG. 4 is a flow chart showing the construction of vector pAV 7

FIG. 5 is a DNA sequence of Example II.

FIG. 6 is a DNA sequence of Example II.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

One aspect of this invention relates to a novel replicon whose physical map is depicted in FIG. 1. As shown in FIG. 1, the replicon includes two essential features. One essential feature is a structural gene which codes for the production of one or more polypeptides comprised of repeating units of one or more peptide sequences. Suitable genes may vary widely and depend on the desired polypeptide. Illustrative of useful genes are those coding for the production of naturally occurring materials or their synthetic analogs. For example, useful genes are those coding for naturally-occurring fibrous or film forming proteins such as collagen, elastin, insect salivary gland silk protein, silk fibroin, troponin C, tropomyosin, and the like and their synthetic analogs such as poly(Gly-Pro-Pro)$_n$, poly(Pro-Gly-Pro)$_n$, poly(Pro-Pro-Gly)$_n$, poly(Val-Pro-Gly-Val-Gly)$_n$, poly(Gly-Ala-Gly-Ala-Gly-Ser)$_n$ and the like where n is an integer of from about 2 to about 200, and preferably from about 15 to about 100.

Similarly, still other useful genes are those which code for naturally-occurring adhesives such as insect salivary gland adhesive protein, bioadhesive proteins from marine crustaceans such as Mytilis edulis, M. californianus, and Geukensia demissa, trematode egg shell dopa proteins, and the like and their synthetic analogs and naturally-occurring architectural proteins such as egg shell proteins, keratin, insect cuticle proteins, and the like, and their synthetic analogs.

Also useful in the practice of this invention are genes coding for the production of synthetic polypeptides such as:

poly-(Ala-Lys-Pro-Ser-Tyr-Pro-Pro-Thr-Tyr-Lys)$_n$;

poly-(Ala-Lys-Pro-Ser-Tyr-4-Hyp-4-Hyp-Thr-Tyr-Lys)$_n$;

poly-(Ala-Lys-Pro-Ser-Tyr-4-Hyp-4-Hyp-Thr-Tyr-Lys)$_n$;

poly-(Ala-Lys-Pro-Ser-Phe-4-Hyp-4-Hyp-Thr-Tyr-Lys)$_n$;

poly-(Ala-Pro-Ser-Tyr-4-Hyp-4-Hyp-Thr-Tyr-Lys)$_n$;

poly-(Ala-Lys-Pro-Ser-Tyr-4-Hyp-4-Hyp-Thr-Try-Lys)$_n$;
poly-(Ala-Lys-Pro-Ser-Tyr-Pro-Pro-Thr-Tyr-Lys)$_n$;
poly-(Lys-Pro-Ser-Tyr-4Hyp-4Hyp-Thr-Tyr-Lys-Ala)$_n$;
poly-(Ala-Lys-Pro/Hyp-Ser/Thr-Tyr/Dopa-Pro/Hyp-Ser/Thr-Tyr/Dopa-Lys$_n$; poly-(Gly-X-Y)$_n$; poly(Gly-Pro-X)$_n$;
poly-(Gly-X-Pro)$_n$;
poly(X-Pro-Gly-Y-Gly)$_n$;poly-(X-Pro-Gly-Gly)$_n$;
poly(X-Pro-Gly-Val-Gly-Y);
poly-(Ala)$_4$-(Lys)-(Ala)$_2$-Lys$_2$-(Phe/Try)-Gly-Ala)$_n$;
poly-(Ala-Gly)$_n$; poly-(Ala)$_2$-Lys-(Ala)$_3$-Lys(Ala$_2$))$_n$; poly-(Gly-Ala-Gly-Ala-Gly-Ser)$_n$;
poly-(Ala-Lys-Pro-Ser-Try-Pro-Pro-Thr-Tyr-Lys)$_n$;
poly-(Pro-Leu-Gly)$_n$; poly-(Ala-Gly-Gl)$_n$;
poly-(Val-Pro-Gly-Val-Gly)$_n$; poly-(Ser-Gly-Gly)$_n$;
poly-(Pro-Phe-Gly)$_n$; poly-(Pro-Lys-Gly)$_n$;
poly(Lys-Gly-Gly)$_n$; poly-(Pro-Gly-Gly)$_n$;
poly-(Pro-Pro-Gly)$_n$; poly-(Ala-Phe-Gly)$_n$;
poly-(Lys-Clu-Gly)$_n$; poly-(Ala-Gly-Gly-Gly)$_n$;
poly-(Pro-Leu-Gly-Gly)$_n$; poly-(Pro-Gly-Pro-Gly)$_n$; and
poly-(Lys-Glu-Lys-Glu)$_n$;

where the amino acids are listed by standard three letter code, "Hyp" is hydroxyproline, "4-Hyp" is 4-hydroxyproline, "Dopa" is 3,4-dihyroxyphenyl alanine, n is equal to or greater than 1, preferably from 1 to about 1000, more preferably from 1 to abut 500 and most preferably from 1 to about 150, and X and Y are the same or different and each is a natural or non-natural amino acid and the nomenclature, X/Y, indicates that either X or Y can be present in the copolymer chain at the sequential position indicated.

In the preferred embodiments of the invention, the gene codes for polypeptides which comprise one or more recurring monomeric units derived from lysine and/or glycine. In these preferred embodiments, the remaining recurring monomeric units can be derived from any of the other amino acids. In the particularly preferred embodiments of the invention, the gene codes for oligomeric and polymeric polypeptides which may also

include one or more recurring units derived from tyrosine or other hydroxy substituted amino acids such as hydroxyproline, hydroxylysine and the like. Amongst these particularly preferred embodiments, most preferred are those embodiments in which the gene codes for the production of precursor polypeptides of naturally occurring bioadhesives such as those including from about 100 to about 1500 amino acid residue sequences consisting of about 20% to about 40% of proline residue, about 10% to about 40% of lysine residue, about 10% to abut 40% of tyrosine residue and 0 to about 40% of amino acid residues other than proline, lysine and tyrosine. Preferably, the precursor protein of the bioadhesive is made of a repeating decapeptide which contains about 30% of proline residue, 20% of lysine residue and 20% of tyrosine residue which is more preferably of the formula:

ALA/LYS-PBD/HYB-SER/THR-TYP/DOPA-
PRP/HYP/PRO/HYP-SER/THR-TYP/DOPA-LYS

which is described in more detail in U.S. Patent No. 4,585,585.

Genes useful in the practice of this invention can be obtained from natural sources or synthesized in accordance with known procedures. For example, useful genes can be synthesized using the procedures of European Patent Application Publication No. 0 154 576, PCT WO 88/03533, PCT WO 87/03369 and PCT WO 87/02822.

As an alternative source of genes for use in the practice of this invention are natural genes or gene fragments or complementary DNA copies of all or a portion of a natural gene in the form of double-stranded DNA fragments which can be isolated by techniques well known in the art of molecular cloning. Illustrative of natural genes or gene fragments which are useful in the practice of this invention are those which code for part or all of any form or isolate of the proteins collagen, elastin, keratin, troponin C, any other intermediate filament

protein (c. E. Lazarides, Nature 283, 249-256 (1980)) or silk fibroin and which includes most or all of an amino acid sequence which exhibits some degree of repetitiveness within the protein sequence. The degree of repetitiveness can be judged by DNA or protein sequence homology using various theoretical techniques in peptide biology. See, for example, S.B. Needleman and C.D. Wunsch, Journal ofMolecular Biology 48, 443-453 (1970), A.D. McLachlan, Journal of Molecular Biology 61, 409-424 (1971), and D. Eisenberg et al., Proc. Natl. Acad. Sci. (U.S.A.) 81, 140-144 (1984). Exemplary of other naturally occurring DNA for use as genes are those resulting from reverse transcription and DNA strand copying from messenger RNA by an appropriate reverse transcription process and DNA strand copying process wherein the messenger RNA is transcribed from gene coding for proteins such as collagen, elastin, keratin, troponin C, any other intermediate filament, or silk fibroin. These natural DNA fragments will preferably be prepared for isolation using a restriction enzyme which leaves cohesive termini on the natural DNA fragments compatible with the cohesive termini of the plasmid vector selected for use. Alternatively, the ends of any natural DNA fragments may be adapted or modified with an appropriate DNA linker or linkers which subsequent to attachment to the natural DNA fragments can either be uniquely cleaved with one or more restriction enzymes to reveal or intrinsically has one or more cohesive termini compatible with the cohesive termini of the cleaved plasmid vector.

As another essential feature, the replicon of this invention includes an effective expression system. As used herein, an "effective expression system" is a system which on transformation of a microbial organism by the replicon of this invention is capable of expressing the gene such yield of the desired polypeptide is equal to or greater than about 10% by weight based on the total weight of cellular protein. In the more preferred embodiments of the invention yields as equal to or greater than about 30%

by weight on the aforementioned basis. In the particularly preferred embodiments of the invention, the expression system will be selected such that the yield of the desired polypeptide is equal to or greater than about 40% by weight on the aforementioned basis and on the most preferred embodiments the yield is from abut 40 to about 60% by weight on the aforementioned basis.

The expression system includes three essential components. One essential component is a transcriptional and translational regulatory region upstream 5' of the structural gene and in reading frame therewith. This region may be created using conventional procedures. For example, this region may be created byemploying a fusion protein, where the subject structural gene is inserted into a different structural gene down stream from its initiation codon and in reading fram with the initiation codon. Various transcriptional and translations initiation regions are available from a wide variety of genes for use in expression host, so that these transcritptional and translationa initiation regions may be joined to the subject structural gene to provide for transcription and translation initiation of the subject structural gene. Preferred ribosome binding regions is a Shine-Delgardo region.

Other essential components include an inducible transcription initiation region or a promoter sequence. In the preferred embodiments, the promoter sequence is an inducible class III promoter sequence or an inducible transcription initiation region. Of particular interest is the use of an inducible transcription initiation region. In this manner, the host strain may be grown to high density prior to significant expression of the desired product. Providing for inducible transcription is particularly useful where the peptide is retained in the cellular host rather than secreted by the host.

A number of inducible transciption initiation regions exist or can be employed in particular situations as for example gene 10 of the T7 expression system. The inducible regions may be controlled by a particular chemical, such as isopropyl thiogalactoside (IPTG) for inducing the beta-galactosidase gene. Other inducible regions include lambda left and right promoters; various amino acid polycistrons, e.g., histidine and tryptophan; temperature sensitive promoters; and regulatory genes, e.g., $cI^{ts}857$.

An alternative system which may be employed with advanges is use of a combination of transcription initiation regions. A first transcription initiation region which regulates the expression of the desired gene but which is not function in the expression host by failing to be fuctional with the endogensous RNA polymerase is employed. A second transcription initiation region, such as an inducible region, can then be employed to regulate the expression of an RNA polymerase with whichthe first transcription initiation region is functional. In this manner expression only occurs upon activation of the regulatory region controlling the expression of the exogenous RNA polymerase. In the subject application, this system is illustrated with the T7 phage transcription initiation region, specifically the initiation regions of genes 9 and 10 of T7 phage.

An alternative system relies on the use of mutants which undergo a developmental change based on a change in the environment, such as a lack of a nutrient, temperature, osotic pressure, salinity, or the like. Illustrative of this system, strains of B. subtilis can be obtained which are incapable of sporulation but which can produce those components which initiate expression of products involved with sporulation. Therefore, by a change in the condition of the medium, a transcription initiation region associated with sporulation will be activated. In this situation, the host provides the necessary inducing agent or activator to initiate expression.

Various other techniques exist for providing for inducible regulation of transcription and translation of a gene in a particular host can be used.

As a third essential component, the replicon includes an initiation codon.

The expression system include various optional components. For example, the expression system may include region for production of RNA polymerase. If not present in the system or in the replicon, such a region may be introduced into the transformed microbial orgamism by several suitable procedures.

There are several acceptable ways to provide a source of RNA polymerase. The RNA polymerase gene can reside on the host chromosome, it can be introduced on a bacteriophage, or it can be carried on a plasmid. The plasmid can be the expression plasmid itself or can be a different plasmid. The expression of the RNA polymerase can be controlled by introducing it on a bacteriophage asa consequence of infection, or by regulated expression systems well nonw to those skilled in the art. Examples are control systems provided by the lactose or tryptophan operon based on the pL promoter (Coning Vectors, P.H. Pouwels, B.E. Enger-Valk and W.J. Brammar, 1989, Elsevier Science Publishing, New York, New York.

The expression construct may optionally include transcriptional and translational termination regulatory region upstream 3' of the structural gene. This region may be created using conventional procedures. For example, this region may be created by employing a fusion protein, where the subject structural gene is inserted into a different structural gene downstream from its initiation codon and in reading frame with the initiation codon. A variety of termiantion regions are available which may be from the same gene as the transcriptional initiation region or from a different gene.

Preferred expression systems for use in the practice of this invention are the T7, T3, SP6 and gh-1 expession systems. See J.F. Klement et al, Gene Anal. Techn., 3:39-66, (1986) and A.H. Rosenberg et al, Gene, 56:129-135 (1987). The most preferred expression system is T7.

Preferred replicons of this invention are pAG9 and PAG11 whose partial genetic maps are depicted in Figure 2. PAG9 includes a T7 Class III promoter sequence downstream of which is a Shine-Delgardo (SD) rebsome binding site. Downstream of this system is a gene which codes for polypeptide containing repeating decapeptide sequences, (X_1). In PAG9, the gene consists of 600 bp formed from twenty sequences each containing thirty base pairs. The gene is bound by a NheI and Nde I restriction recognition sites and downstream of which is a BAMHI restriction recognition site and a transcription termination sequence (TO). pAG16 is substantially the same as pAG9 except that the gene which codes for the desired decapeptide consists of 600 base pairs formed from five, sequences of 120 base pairs.

The replicon of this invention can be formed from a plasmid vector using conventional techniques. Suitable vectors are those which contain T7 expression system andsuitable restriction recognition sites positioned such that on insertion of the gene, the gene can be expressed by the T7 expression system to provide the desired polypeptide. Suitable vectors are also those which can be cleaved to provide an intact replicator locus and system where the linear segment has ligatable termini or is capable of being modified to introduce ligateable termini. Of particular interest are those plasmids which have a pheno typical property, which allow for ready separation of transformants from the parent microorganism. The plasmid vector will be capable of replicating in a microorganism, particularly a bacterium which is susceptible to transformation. Various unicellular microorganisms can be transformed, such as bacteria, fungi and algae. That is, those unicellular

organisms which are capable of being grown in cultures of fermentation.

A wide variety of plasmid vectors may be employed of greatly varying molecular weight. Normally, the plasmid vectors employed will have molecular weights in the range of about $1X10^6$ to $50X10^6$d, more usually from about 1 to $20X10^6$d, and preferably, from about 1 to $10X10^6$d. The desirable plasmid size is determined by a number of factors. First, the plasmid vector must be able to accommodate a replicator locus and one or more genes that are capable of allowing replication of the plasmid. Secondly, the plasmid vector should be of a size which provides for a reasonable probability of recircularization with the foreign gene(s) to form the recombinant plasmid chimera. Desirably, a restriction enzyme should be available, which will cleave the plasmid vector without inactivating the replicator locus and system associated with the replicator locus. Also, means must be provided fo providing ligatable termini for the plasmid vector, which are complementary to the termini of the foreign gene(s) to allow fusion of the two DNA segments.

Another consideration for the recombinant plasmid vector is that it be compatible with the bacterium to betransformed. Therefore, the original plasmid vector will preferably be derived from a member of the family to which the bacterium belongs.

The original plasmid should desirably have a phenotypical property which allows for the separation of transformant bacteria from patent bacteria. Particularly useful is a gene, which provides for survival selection. Survival selection can be achieved by providing resistance to a growth inhibiting substance or providing a growth factor capability to a bacterium deficient in such capability.

Conveniently, genes are available, which provide for antibiotic or heavy metal resistance or polypeptide resistance, e.g. colicin. Therefore, by growing the bacteria on a medium containing a bacteriostatic or

bacteriocidal substance, such as an antibiotic, only the transformants having the antibiotic resistance will survive. Illustrative antibiootics include tetracycline, streptomycin, sulfa drugs, such as sulfonamide, kanamycin, neomycin, penicillin, chloramphenicol, or the like.

Growth factors include the synthesis of amino acids, the isomerization of substrates to forms which can be metabolized or the like. By growing the bacteria on a medium which lacks the appropriate growth factor, only the bacteria which have been transformed and have the growth factor capability will survive.

A large number of suitable vectors are commercially available, with others being described in the literature. One preferred plasmid vector for in the practice of this invention pET-3a whose genetic map is depicted in Figure 3.

The gene which code for the desired repeating amino acid sequence can be inserted into a suitable plasmid vector using conventional techniques. Such techniques are well known in the art, and will not be described herein in detail. See for example U.S. Patent No. 4,237,224 and reference cited thereon. The gene will preferably be inserted at a unique site or pair of sites in the plasmid vector that allows perfect base pairing with cohesive termini on the gene. Such insertion may or may not yield a restriction enzyme recognition sequence at any of the junctions between the plasmid vector and the inserted gene. In the preferred embodiments of this invention, such a restriction enzyme recognition sequence is constituted or reconstituted so that the inserted gene may be removed at a later time if desired in other applications of this invention. The site of gene insertion is preferably at a position 3' to the expression system, the plasmid vector which will regulate the production of sufficient amounts of polypeptide from the inserted gene which must be inserted in the correct reading frame and in the proper orientation.

The replicon of this invention can be transformed using conventional techniques known in the art of molecular cloning using an acceptable bacterial host or other suitable microorganism in which the gene in the replicon is capable of being expressed using established techniques, as for example those techniques described in U.S. Patent 4,237,224; T. Maniatis et al., Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, 1983), pp. 249-255; and D. Hanahan, Journal of Molecular Biology, 166, 557-580 (1983) and incorporated herein by reference.

Useful bacterial species may vary widely and include species of the genus Enterobacteriaceae, Salmonella, Bacillaceae, Pneumococcus, Streptococcus, Pseudomonas, Methylomonas, Saccharomyces, and Rhodopseudomonas such as Saccharomyces cerevisiae, Pseudomonas spp, Streptomyces coelicolar, Escherichia coli, Bacillus subtilis and the like. Preferred bacteria are strains of E. coli especially those which are recombinant-deficient in orderto prevent recombination events that may be favored between various segments of the inserted gene which have a substantial degree of internal repetitiveness. Especially preferred strains of E. coli are genotype rec A-, especially MH01 (genotype recA-, Tet^r derivative of strain N99) whose construction is described in the examples below, MH03 (recA-, Tetr derivative of strain N4830 made by P1 transduction from strain N6240 by techniques analogous to those used in the construction of MH01), DC1138 (pro^-, Leu^-, srlR recA301::Tn 10, def^{cI}), DC1139A (same as DC1138 except def Bam H1 H1 cI857), JM109 and DHB9 (F'lac^g_I Z^+Y^+, recA, srl::Tn 10, phoR, phoA, malF, ara leu, lac, galE, galK; derived from MC1000).

After transformation, clonal isolates of transformed bacteria can be screened and selected using conventional techniques as for example screening by hybridization techniques using a radio labelled synthetic oligodeoxynucleotide probe. The screened bacterial colonies can be s lected and isolated once it is

determined that they contain useful plasm d vectors, and can be assayed for expressing the inserted gene as a polypeptide with the desired repeating amino acid sequence.

Subcultures of the most preferred microorganisms of this invention (AS002 (pAG9) and AS002 (pAG16) have has been deposited in the permanent collection of the Northern Regional Research Laboratories, Agricultural Research Services, U.S. Department of Agriculture, Peoria, Ill., USA, under the accession numbers NRRL B-18544 and B-18545. The permanency of the deposits of these cultures and ready accessiblity thereto by the public are afforded throughout the effective life of the patent in the event the patent is granted. Access to the cultures is available during pendency of the application under 37 CFR 1.14 and 35 USC 112. All restrictions on the availability to the public of the deposited cultures will be irrevocably removed upon granting of a patent.

If cloned bacteria are capable of polypeptide expression from the gene in additional bacteria can be grown under fermentation conditions and these bacteria can induced to express the desired polypeptide under conditions which are appropriate for the particular plasmid vector-bacterial host gene expression system being utilized. The desired polypeptide can then be isolated from the bacterial growth medium or from the bacteria using appropriate procedures. Illustrative of useful bacterial growth and bacterial produce harvest procedures are those described in greater detail in European patent application 0131843 which is incorporated herein by reference.

This invention has many uses. For example, the invention can be used to make or create bacteria which produce many useful polypeptide products. Illustrative of such products are analogues to naturally occurring proteins such as collagen, elastin, keratin, protein or glycoprotein elements of think, intermediate or thin filaments in higher organisms, silk fibroin, tropomyosin, troponin C, resilin, eucaryotic egg shell proteins, insect cuticle proteins or other eucaryotic architectural proteins.

The following examples are presented to more particularly illustrate the invention and are not to be construed as limitations thereon.

EXAMPLE I

(A) Construction of a gene cassette encoding the consensus decapeptide for the bioadhesive protein from M. edulis: The following four oligodeoxynucleotides were synthesized by solid-phase synthesis using phosphoramiditte chemistry on an Applied Biosystmes 380B DNA synthesizer:

a. 5'-CCAACCTACAAAGCTAAGCCGTCTTATCCG-3'
b. 5'-GTAGGTTGGCGGATAAGACGGCTTAGCTTT-3'
c. 5'-CCAACCTACAAAGCCAAGGCTTCTTATCCG-3'
d. 5'-GTAGGTTGGCGGATAAGAAGCCTTGGCTTT-3'

Oligodeoxynucleotides a and c encode the decapeptide sequence described by J.H. Waite, J. Biol. Chem. 258, 2911-2915 (1983) except for a substitution of Ala for Pro in the third position of c. These oligeoxynucleotides were used to build a glue decapeptide analog gene cassette containing StyI ends essentially as described in PCT WO US 87/03369. Briefly, one nmol each of a and b were separately phosphorylated with T4 kinase. Separate phosphorylation reactions using [$-^{32}$P] ATP were employed to prepare oligodeoxynucleotides c and d. Phosphorylated oligodeoxynucleotides c and d (20 pmol) were purified from unincorporated radioactivity by passage over a NENSORB column. Oligodeoxynucleotide pairs a and b, and c and d were combined, heated to 70 C for 15 minl, and allowed to cool to 45 C over 3 hrs. in order for each pair to anneal to form short duplex DNA with 9 base 5' overhangs. The annealed c-d pair contains a recognition site for the restriction endonuclease StyI. Four hundred pmol of annealed a-b was added to c-d and the temperature was allowed to cool to room temperature overnight to allow

for formation of long duplex a-b interspersed with c-d segments. Ligase was used to close the nicks between adjacent oligodeoxynucleotides by incubating at room temperature for 8 hrs. The ligase was heat inactivated and the long repetitive DNA so formed was subjected to digestion with the restriction endonuclease StyI overnight at 37 C. The resulting glue decapeptide gene cassettes containing StyI ends were purified by size-exlusion chromatography on a Sepharose 4B column after inactivating the StyI enzyme. The size of these glue cassettes in various fractions was ascertained by polyacrylamide gel electrophoresis of sample aliquots. DNA was recovered from fractions in which most molecules had lengths greater than about 100 bp. Approximately 20 pmol of glue decapeptide analog gene cassette was recovered, and of this about 5 pmol was used in a ligation to 0.5 pmol of the cloning and expression vector pAV7 in 60 ul of reaction buffer. The reaction was incubated in the presence of ligase for 4 hrs. at room temperature and then diluted to 1 ml in TE buffer. Two to 10 ul were used to transform E. coli strain DC1138.

(B) Construction of cloning and expression vector pAV7:

The construction of vector pAV7 is illustrated in Figure 4. Vector pAV7 was derived from plasmid pJL6, which is described in J.A. Lautenberger, D. Court and T.S. Papas, Gene 23, 75-84 (1983) Plasmid pJL6 is an expression vector based on the temperature-inducible leftward promoter of bacteriophage lambda. Plasmid pJL6 was digested with restriction endonucleases PvuII and EcoRV, both of which generate blunt ends. The large DNA fragment was purified and a 29vb synthetic DNA fragment representing a bacteriophage SP6 promoter was inserted via ligation with T4 DNA ligase. The SP6 promoter consisted of the following sequence:

5'-ATTTAGGTGACACTATAGAATAGGGATCC-3'
3'-TAAATCCACTGTGATATCTTATCCCATGG-5'

Neither the PvuII nor EcoRV restriction sites were regenerated. This vector was designated pAV01.

Plasmid pAV01 was then opened at the single AvaI restriction site and made blunt ended with the large (Klenow) fragment of E. coli DNA polymerase I. A 29 bp synthetic DNA fragment encoding a bacteriophage T7 promoter was inserted which contained the following sequence.

5'-TAATACGACTCACTATAGGGAGATCGCGA-3'
3'-ATTATGCTGAGTGATATCCCTCTAGCGCT-5'

The AvaI restriction site was not regenerated. This vector was designated pAV02 The T7 and SP6 promoters are located upstream and downstream of the cloning region, respectively, to allow for in vitro transcription of anyinsert from either direction upon addition of the appropriate RNA polymerase.

Two additional oligodeoxynucleotides were synthesized, 15 and 17 bp long, which upon annealing gave rise to a double-stranded DNA fragment ending in NdeI and HindIII restriction sites. These in turn flanked the 6 bp recognition sequence for restriction endonuclease StyI. The DNA fragment has the following sequence:

StyI
5'-TATGGCCAAGGCTTA-3'
3'-ACCGGTTCCGAATTCGA-5'

The StyI restriction site was chosen as the cloning site for polydecapeptide analog gene cassettes because it is asymmetric anc changes only one codon from proline to alanine in the third position of the decapeptide consensus sequence. The plasmid pAV02 was then simultaneously digested with restriction enzymes NdeI and HindIII. The large fragment was purified and the synthetic DNA fragment containing the StyI cloning sit was inserted with T4 DNA ligase. This vector is approximately 3.2 kilobases long and was designated pAV7.

All recombinant DNA techniques employed are known to those skilled in the are and ar described in T. Maniatis et al., Molecules Cloning (Cold Spring Harbor). All synthetic DNA sequences were made on an Applied Biosystems Model 380B DNA synthesizer.

(C) Cloning and characterization of glue analog gene cassettes: Oligodeoxynucleotide a was phosphorylated using [$-^{32}$P]ATP in order to generate a probe for colony hybridization. Transformed colonies were screened with the probe, and those showing intense hybridization were chosen for isolation of plasmid DNA. Plasmid preparations were digested with the following restriction enzyme pairs to identify the presence and relative size of the polydecapeptide gene cassettes; EcoRI-HindIII and NruI-HindIII. The proper insertion of the glue analog gene cassette into the pAV7 vector was confirmed by demonstrating restriction of the plasmid DNA by StyI restriction endonuclease to liberate the full-sized glue analog gene cassette with StyI ends. Strains containing the plasmid carrying polydecapeptide gene cassettes with StyI ends were archived into the culture collection and contained gene cassettes ranging in size from 120 bp (pAG4) to 600 bp (pAG3).

(D) Expression of putative polydecapeptide from the glue analog gene cassette of pAG3: E. coli strain IG110 was transformed with pAG3 and the vector control pAV7. The utility of strain IG110 for production of repeating peptides is described by Goldberg and Salerno in US Patent Application No. 251,714. Briefly, IG110(pAV7) and IG1110(pAG3) were grown separately to about 10^8 cells/ml in 10 ml LB-ampicillin broth at 30 C. The cells wee then filtered and resuspended in 10 ml of M63 salts including glucose, vitamin B_1 and all amino acids except proline. Each culture was divided into two 5 ml aliquots and each was incubated at 30 C and 41 JC for 10 min. Two uCi [^{14}C]proline/ml of medium was then added to each culture and the incubation continued for 20 min. Cells were

chilled in ice water, then centrifuged, and the cell pellets were resuspended in 0.15 ml of 50 mM tris-HCl, 2% -mercaptoethanol, 0.5% CTAB and 1 mM PSMF. Cells were broken open by sonication for 3 ten-second intervals. Protein concentration was determined by the Bradford assay and 100 g of each sample was analyzed by electrophoresis on a 15% polyacrylamide gel containing 150 parts acrylamide to 1 partbis-acrylamide. The gela dn buffer consisted of 0.9 M acetic acid, 2.5 M urea and 0.01% CTAB. Autoradiography demonstrated that only the IG110 (pAG3) culture at 41 C and not the other 3 cultures showed a highly labeled protein band. Other protein bands from all four cultures were any faintly visible at exposure times which allowed the unique band from the IG110(pAG3) culture to be readily detected.

In another example, IG110(pAV7) and IG110(pAG3) were grown to 10^8 cells/ml in LB broth at 30 C. Then, the cultures were shifted to 41 C and samples were taken at 0.5, 1, 2 and 3 hrs. Cells were recovered, prepared for analysis and analyzed as described above. This time however, proteins were stained with Coomassie brilliant blue to detect the putative polydecapeptide. A unique protein band was visible in all IG110(pAG3) lanes that could not be detected in the IG110(pAF7) lanes. This protein band co-migrates with the novel protein band has a mobility relative to histone H1 that is consistent with its theoretical size of 25,000 daltons. The intense radio labeling is an indication of the relatively high (30%) proline content of polydecapeptide.

EXAMPLE II

A gene was designed in such a way that the 10 amino acid core unit was repeated four times with maximum diversity in the DNA sequence of 120 base pairs (See Figure 5). This 120 base pair region is flanked by recognition sites for the restriction enzyme StyI to be used in cloning of the gene. Two oligonucleotides having

the sequences composition A and B (See Figure 2) were synthesized on an Applied Biosystems model 380 DNA synthesizer. The two oligonucleotides were complementary at their 3' ends for a length of 15 base pairs. The oligonucleotides are derived from a controlled pore glassmatrix and partially purified by use of an OPC column (Applied Biosystems, Foster City, CA) according to manufacturers directions. Then further purified by excision of the full length product by polyacrylamide gel electrophoresis in the presence of 8 M urea. The oligonucleotides were recovered from the gel slices by either a crush and soak or electro elution method. Each oligonucleotide was passed over a G-25 size exclusion column to change buffer (to 1/10 TE) and concentrated 10X. The two oligonucleotides are then combined in a microfuse tube (10 mg of each is ample) in a small volume (25-100 1), heated to70 C in a glass beaker containing 700 mls of water, and allowed to cool to 37 C forming the annealed oligonucleotides with the hydrogen bonds of the complementary base pairs forming a stable double-stranded 3' overlap region. The double stranded 3' overlap region is then extended towards each end using 40 units of sequences (modified T4 DNA polymerase) 40 mM This HCl pH8.0, 10 mM $MqCl_2$, 5 mM DTT, 50 mM NaCl, 1 mM dNTPs, and 50 ug/ml BSA. The above reaction mixture was allowed to incubate at 37 C for 4 hours. The sequenase enzyme was then heat inactivated. The reaction was phenol/chloroform extracted and the resulting DNA was ethanol precipitated. This DNA segment was digested with the restriction endonuclease StyI forming a gene cassette for cloning.

EXAMPLE III

(A) Preparation of AS002(pAG9)

AS002(pAG9) was made by first removing the glue decapeptide analog gene cassette from pAG3 as a NdeI-BamHI DNA fragment. The translation initiation region of the

glue decapeptide analog gene cassette comprises sequences from the latter as well as the expression vector itself. The DNA sequence of this region is indicated below:

5'-GGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGAT
ATACATATGGCCAAGGCTTCTTATCCG-3'

The plain uppercase letters refer to the vector DNA and the italicized letters refer to sequence derived from the gene cassette. The underlined sequence indicates the initiation codon for Met. This same nomenclature is used for all other examples as well.

This fragment was then ligated to the large NdeI-BamHI DNA fragment of pET3a to give pAG9. pAG9 was introduced into the production strain AS002 by transformation. AS002 was derived by moving the (srlR-recA)306::Tn10 allele from DC1138 to the production strain BL21/DE3(pLysS) by P1 transduction. All these manipulations were conduct with standard techniques well known to those skilled in the art (Current Protocols in Molecular Biology, F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, K. Struhl, eds., 1989, John Wiley & Sons, New York, New York; Experiments in Molecular Genetics, J.H. Miller, 1972, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).

The host BL21/DE3(pLysS) and the expression vector pET3a together represent one of several embodiments of the T7 expression system (A.H. Rosenberg et al., Gene 56:125-135, 1987). The system functions to produce protein (or in our case polypeptide repeats) in the following manner. The host BL21 contains the lambdoid phase DE3 on its chromosome. DE3 is a chimeric phase containing the gene for the bacteriophage T7 RNA polymrase under the control of the promoter and operator sequences of the lactose operon. By adding a B-D-galactopyranoside to the bacterial culture such as IPTG, transcription from the lactose operon promoter can be induced and this results in the production of T7 RNA polymerase from the

gene. However, in the absence of inducer a small basal level of T7 RNA polymerase is produced and even this amount of T7 RNA polymerase can cause cell death in the presence of a toxic gene residing on an appropriate expression plasmid. Therefore, BL21/DE3(pLysS) contains the plasmid pLsyS which provides for a low level synthesis of T7 lysozyme. T7 lysozyme serves two purposed in the cell. Most importantly, T7 lysozye acts to inhibit the function of T7 RNA polymerase. Secondly, it aids in purification of product proteins by breaking down prat of the cell wall during the initial steps in the isolation. The expression vector pET3a contains the following components from gene 10 of the bacteriophage T7; the strong Class III promoter, the translation initiation region, and the 5' region of the structural gene. In pAG9, the glue decapeptide analog gene cassette replaces the structural gene. T7 RNA polymerase makes messenger RNA to glue decapaptide analog gene cassette using the strong Class III T7 RNA polymerase promoter of gene 10. This mRNA is translated into polydecapeptide using the Shine-Delgarno and surrounding sequences of the translation initiation region of gene 10.

(B) EXPRESSION OF AS002(pAG9) to Make Polypeptide

AS002(pAG9) is used to make polydecapeptide by growing the culture in LB broth until the culture reaches an optical density of at least 0.5-1.0 at 600 nm. IPTH is then added to the culture to a final concentration of 0.4 mM and the cells are incubated an additional 3 hours before harvesting. Ampicillin is added to a level high enough to ensure that greater than 97% of the cells retain pAG9. For cultures up to one liter, 100 g/ml has been found acceptable. Yield was 40 to 60% of cellular protein and the purified dicapeptide was not fused and had a molecular weight of about 25,000 where the molecular weight distribution (Mw/Mn) was about 1.

EXAMPLE IV

(A) Preparation of AS002(pAG11)

The strain is comprised of the same microbial host as AS002(pAG9) but instead contains the plasmid pAG11. pAG11 was constructed so as to produce a fusion protein containing polydecapeptide. The first 11 amino-terminal residues of the bacteriophage T7 gene 10 protein are fused to the amino-terminal side of polydecapeptide by way of Gly-Ser dipeptide as shown:

met-ala-ser-met-thr-gly-gly-gly-gly-met-gly-arg-gly-ser-met-ala-lys-ala-ser-tyr-pro

The translation initiation region of the expression vector remains unchanged in this construct. The construct was prepared as follows. The glue decapeptide analog gene cassette of the expression vector pAG3 was removed as a NdeI-BamHI DNA fragment. The ens of the NdeI-BamHI DNA fragment were filled in by treatment with the Klenow fragment of DNA polymerase I. The expression vector pET3a was linearized by treatment with BanHI and the ends were als filled by treatment with the Klenow fragment of DNA polymerase I. The glue decapeptide analog gene cassette was then ligated into the linearized pET31 vector, transformed into strain AG1, and pAG11 was identified by restriction mapping. pAG11 was then transformed into strain AS002 to give the final strain AS002(pAG11). The DNA sequence of the region around the junction of the 5' side of the glue decapeptide analog gene cassette and the vector is indicated below:

5'-GGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGAT
ATACATATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGCGGATCTATGGCCAAGGCT
TCTTATCCG-3'

(B) Expression Use of AS002(pAG11) to Make Polydecopertide

This strain is used in essentially like manner to that already described above for AS002(pAG9). Cells are grown until the culture reaches an optical density of at least 0.5-1.0 at 600 nm. IPTG is then added to the culture to a final concentration of mM and the cells are incubated an additional 3 hours before harvesting. Ampicillin is added to a level high enough to ensure that greater than 97% of the cells retain pAG9. for cultures up to one liter, 100 g/ml has been found acceptable. This strain produces fusion dicapeptide at a level 50% of total cell protein.

EXAMPLE V

(A) Preparation of AS002(pAG12)

This strain is also comprised of the same microbial as AS002 (pAG12) host but instead contains the plasmid pAG12. pAG12 too was constructed so as to produce a fusion protein containing polydecapeptide and is the same as pAG11 except fo the length of the fusion. In this construct, the first 260 amino acids of the T7 bacteriophage gene 10 protein are fused through the Gly-Ser dipeptide to the amino-terminal side of polydecapeptide. pAG12 was prepared in the same manner as pAG11 except the expression vector digested the BamHI was pET3xa rather than pET3a. pET3xa is the same as pET3xa execpt that the BamHI restriction site in pETsa is at codon 11 of gene 10 while in pET3xa the BamHI restriction site is at codon 260 of gene 10.

(B) Use of AS002(pAG12) to Make Polydecapeptide

AS002(pAG12) was used in essentially like manner to that already described above for AS002(pAG9). Cells are

grown until the culture reaches an optical density of at least 0.5-1.0 at 600 nm. IPTG is then added to the culture to a final concentration of 0.4 mM and the cells are incubated an additional 3 hours before harvesting. Ampicillin is added to a level high enough to ensure that greater than 97% of the cells retain pAG9. For cultures up to one liter, 100 g/ml has been found acceptable.

This strain produces the fusion polydecapeptide at a level of about 13% of total cell protein.

EXAMPLE VI

(A) Preparation of AS002(pAG16)

This strain utilizes the same microbal host as described above but contains the plasmid pAG16. This plasmid (derived fromp ET3a) harbors a glue decapeptide analog gene cassette approximately 600 bp-long which consists of fiber consecutive repeats of the 120 bp long diversified glue cassette described in Example II. The glue cassette in pAG16 is similar in length to the highly repetitive glue cassette of pAG9.

Plasmid pAG16 was constructed as follows. The 120 bp-long diversified glue cassette was inserted into the StyI cloning site of pAF7 giving rise to pAG13. In order to generate the 600 pbp-long cassette fragment was isolated. The preparation was then ligated to itself and samples were taken at 30, 60, 90 and 120 minutes. The samples were pooled and the ligation products analyzed. The DNA fragment with a length of 600 bp was isolated from a low-melting-point agarose gel and reinserted into StyI-digested pAV7 DNA. About 50 colonies were screened. One clone contained an insert of the correct length and was designated pAG15. The host strain for these constructions was E. coli DC1138. The 600 bp diversified glue cassette eas then moved into expression vector pET3a. Plasmid pAG15 was simultaneously digested with restriction endonucleases NdeI+BamHI. The

cassette-containing DNA fragment ws isolated and ligated into NdeI+BamHI-digested pET3a DNA, giving rise to pAG16. The junction sequence is illustrated below:

5'-GGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGTA
TACATATGGCCAAGGCCAGCTATCCCCCAACGTATAAGGCTAAACCGAG-3'

This cloning step was performed in E. coli HB101. Finally, pAG16 was transformed into the expression host AS002 using the Hanahan procedure.

(B) Use of AS002(pAG16) to Make Polypdecapeptide

AS002(pAG16) is used in essentially like manner to that already described above for AS002(pAG9). Cells are grown until the culture reaches an optical density of at least 0.5-1.0 at 600 nm. IPTG is then added to the culture to a final concentration of 0.4 mM and the cells are incubated an additional 3 hours before harvesting. Ampicillin is added to a level high enough to ensure that greater than 97% of the cells retain pAG9. For cultures up to one liter, 100 g/ml has been found acceptable. The glue decapeptide produced from pAG16 is similar in size to the one produced from pAG9 as analyzed by gel electrophoresis. The yield of polydecapeptide in AS002(pAG16) is approximately 20% of total cellular protein.

WHAT IS CLAIMED IS:

1. A replicon capable of expressing a polypeptide comprising one or more repeating peptide sequences, said replicon comprising in sequence:

an expression system comprising a ribsome binding site, a promoter and an initiation codon; and

one or more structural genes which code for said polypeptide downstream of said expression system, said gene being controllable by said system, whereby said genes are expressible to form said polypeptide when said replicon is cloned into a suitable host microbial organism such that the yield of said polypeptide is equal to or greater than about 10% by weight based on the total weight of cellular protein.

2. A replicon according to claim 1 wherein said yield is equal to or greater than about 30% by weight.

3. A replicon according to claim 2 wherein said yield is equal to or greater than about 40% by weight.

4. A replicon according to claim 1 selected from the group consisting of replicons cloned into E. coli NRRL B-18544 and E. coli B-18545.

5. A novel strain of bacterial host organisms comprising the replicon of claim 1.

6. A strain according to claim 5 comprising the replicons of E. coli NRRL B-18544 or E. coli B-18545.

7. A strain according to claim 5 wherein said bacterial host organism is Escherichia coli.

8. The E. coli strain NRRL B-18544.

9. The E. coli strain NRRL B-18545.

10. A method for producing a polypeptide which comprises culturing the bacterial host organism of claim 5.

FIG. 1

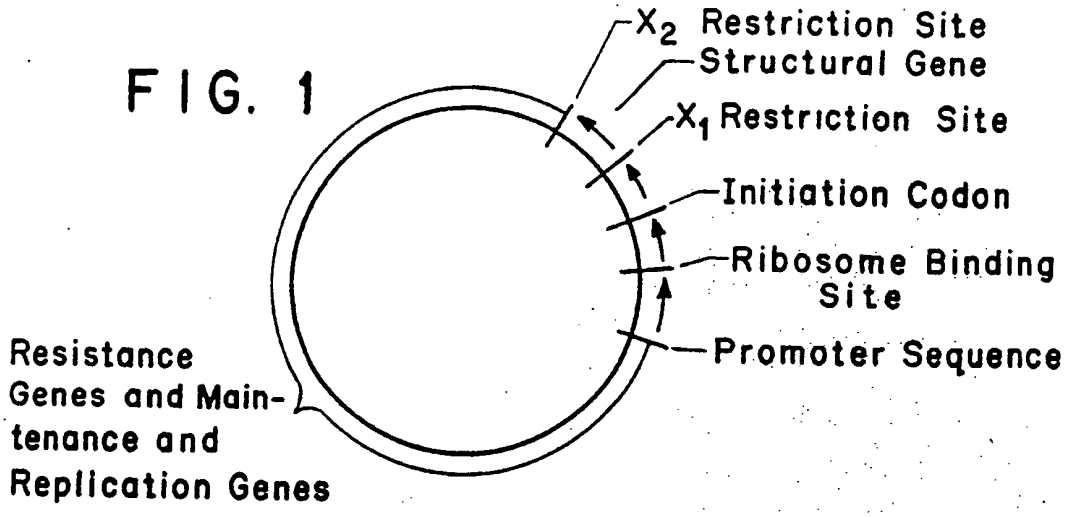
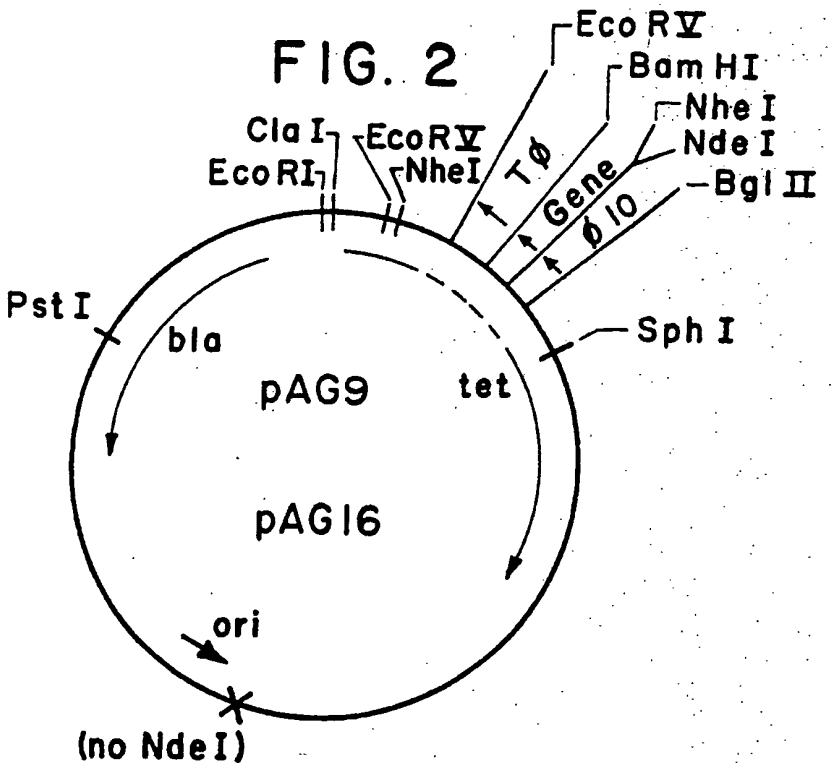

X_2 Restriction Site
Structural Gene
X_1 Restriction Site
Initiation Codon
Ribosome Binding Site
Promoter Sequence
Resistance Genes and Maintenance and Replication Genes

FIG. 2


EcoRV
BamHI
NheI
NdeI
BglII
ClaI
EcoRI
EcoRV
NheI
Tø
Gene
ø10
PstI
bla
pAG9
tet
SphI
pAG16
ori
(no NdeI)

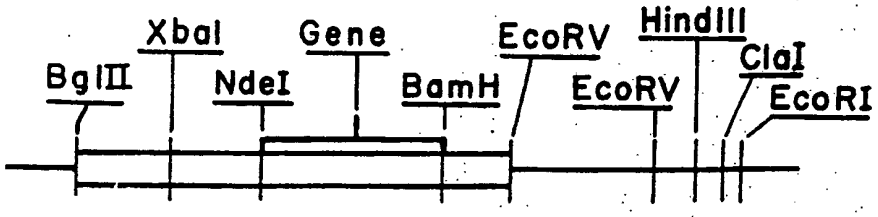
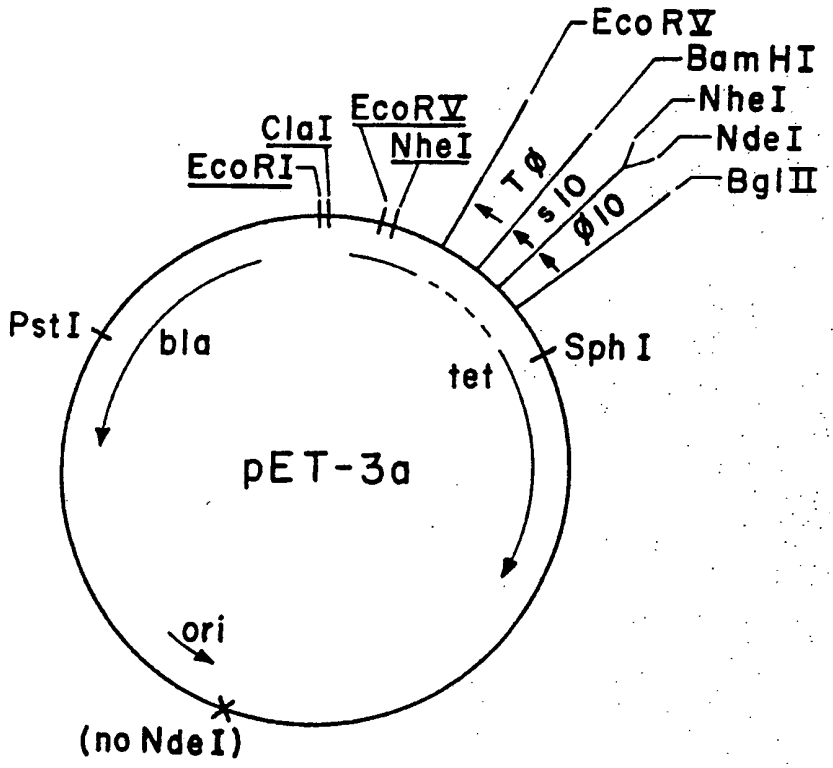
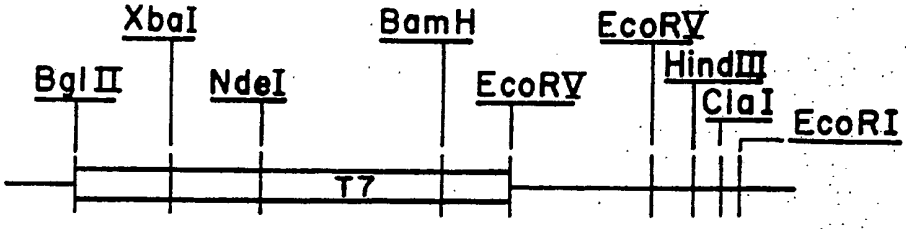

XbaI
Gene
EcoRV
HindIII
BglII
NdeI
BamH
EcoRV
ClaI
EcoRI

FIG. 3





NdeI BamHI

CAT ATG GCT AGC ATG ACT GGT GGA CAG CAA ATG GGT CGC GGA TCC
M A S M T G G Q Q M G R G F

pET3a

FIG. 4

Construction of Vector pAV7

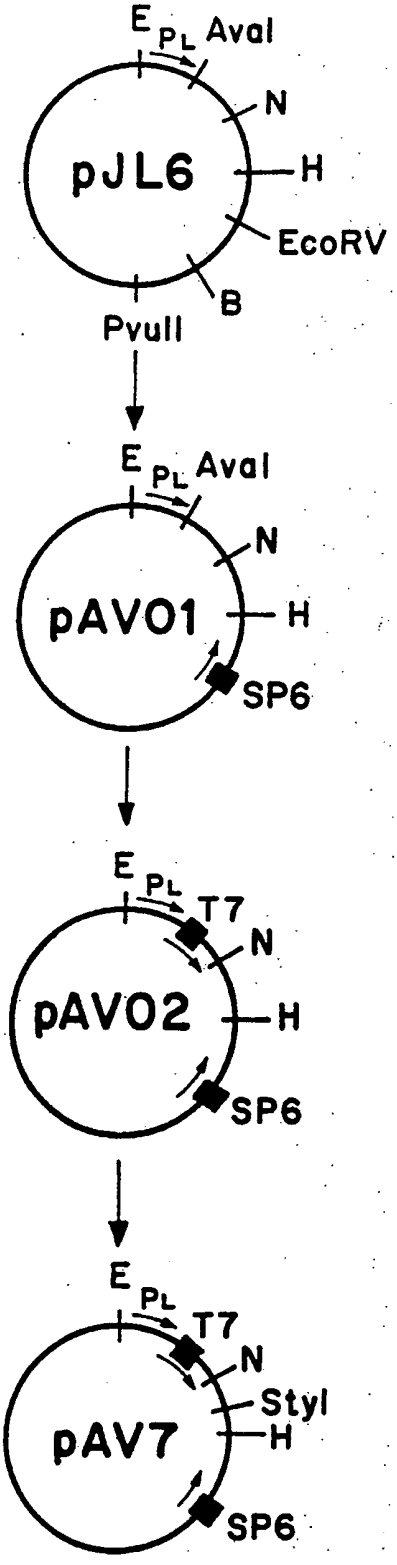


FIG. 5

GLUE 126 BP
OLIGONUCLEOTIDE

```
  1 GCCAAGGCCA GCTATCCCCC AACGTACAAG GCTAAACCGA GTTACCCTCC CACATATAAA
 61 GCAAAACCAT CGTATCCGCC TACCTATAAG GCGAAGCCCT CATACCCACC GACTTACAAA
121 GCCAAG
```

Sequence displayed from position 1 to end (position 126)
Sequence numbered from position 1

FIG. 6

Composition B

StyI

```
  1 GAGTTGACCTACGTAATGCAGCC AAG GCC AGC TAT CCC CCA ACG TAT AAG GCT
                          1▶ Lys Ala Ser Tyr Pro Pro Thr Tyr Lys Ala
 54 AAA CCG AGT TAC CCT CCC ACA TAC AAA GCA AAA CCA TC 3'
```

Composition A

```
                                  3'- G TTT CGT TTT GGT AGC ATA GGC
 11▶ Lys Pro Ser Tyr Pro Pro Thr Tyr Lys Ala Lys Pro Ser Tyr Pro
                                                          StyI
 99
    GGA TGG ATA TTT CGC TTC GGG AGT ATG GGT GGC TGA ATG TTC CGG
 26▶ Pro Thr Tyr Lys Ala Lys Pro Ser Tyr Pro Pro Thr Tyr Lys Ala
144
    TTCCATGCAACTCCAGTAGTTTG -5'
```

INTERNATIONAL SEARCH REPORT

International Application No PCT/US 90/06354

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC[5]: C 12 N 15/70, C 07 K 13/00, C 07 K 7/00, C 12 P 21/02, C 12 N 15/62, C 12 N 1/21, //(C12 N 1/21; C 12 R 1:19)

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
IPC[5]	C 12 N, C 07 K

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category [10]	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
X	EP, A, 0173280 (BIOTECHNOLOGY GENERAL) 5 March 1986 see page 123, example 20; claims 130, 131,137	1-3,5,10
Y		10

X	Gene, volume 15, 1981, Elsevier/North-holland biomedical press, E. Remaut et al.: "Plasmid vectors for high-efficiency expression controlled by the P_L promoter of coliphage lambda", pages 81-93 see the whole document	1-3,5
Y		10

Y	GB, A, 2162190 (PA CONSULTING SERVICES LTD) 29 January 1986 see the whole document	10

---

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
6th March 1991	[illegible] 1991
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	[illegible]

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO. US 9006354 SA 42437

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 25/03/91
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
EP-A- 0173280	05-03-86	US-A-	4742004	03-05-88
		AU-B-	603331	15-11-90
		AU-A-	4667785	06-03-86
		AU-A-	6019890	08-11-90
		JP-A-	61111693	29-05-86
GB-A- 2162190	29-01-86	EP-A-	0230702	05-08-87